arXiv:physics/0701167v1 15 Jan 2007

Predicting hurricane numbers from Sea Surface Temperature: closed form expressions for the mean, variance and standard error of the number of hurricanes

Stephen Jewson (RMS)*

January 16, 2007

Abstract

One way to predict hurricane numbers would be to predict sea surface temperature, and then predict hurricane numbers as a function of the predicted sea surface temperature. For certain parametric models for sea surface temperature and the relationship between sea surface temperature and hurricane numbers, closed-form solutions exist for the mean and the variance of the number of predicted hurricanes, and for the standard error on the mean. We derive a number of such expressions.

1 Introduction

One way to try and predict future hurricane numbers is to predict sea surface temperatures (SST), and then to predict hurricane numbers as a function of the predicted SSTs. If both the prediction of SSTs and the model that relates SSTs to hurricane numbers are probabilistic then the resulting probabilistic prediction of hurricane numbers can, in general, only be derived using numerical methods. However, in certain cases a lot of information about the predicted distribution of hurricane numbers can be derived analytically. This includes estimates for the mean number of hurricanes, the variance of the number of hurricanes, the standard error on the estimate of the mean number of hurricanes and the linear sensitivity of the mean number of hurricanes to changes in the mean and variance of SST.
In this article we derive a number of such relations, for the following two cases:

1. we predict the SST distribution, and then use these predicted SSTs to predict either basin hurricane numbers or landfalling hurricane numbers.

2. we predict the SST distribution, use this to predict the distribution of basin hurricane numbers, and then use a further relationship to predict landfalling hurricane numbers from basin hurricane numbers.

The assumptions we make to render this problem tractable are as follows:

- the sea surface temperature is taken as normally distributed (although in several of our derivations we relax this assumption and consider a completely general distribution of SST with known mean and variance)

- the relationship between sea surface temperature and hurricane numbers is taken as either (a) linear and normally distributed, (b) linear and poisson distributed, or (c) exponential and poisson distributed.

- the relationship between hurricane numbers in the basin and hurricane numbers at landfall is taken to be linear and poisson distributed

This article proceeds as follows. In section 2 we discuss our choice of models and assumptions. In section 3 we discuss the types of statistical models we will use, and the terminology we will use to describe them. In section 4 we describe our notation and what we need from the SST forecasts.
In section 5 we derive expressions for aspects of the predicted distribution of hurricane numbers in the case where the relationship between SST and hurricane numbers is linear and normally distributed. In

* *Correspondence email*: stephen.jewson@rms.com

section 6 we derive expressions for aspects of the predicted distribution of hurricane numbers in the case where the relationship between SST and hurricane numbers is linear and *poisson* distributed. In section 7 we derive expressions for aspects of the predicted distribution of hurricane numbers in the case where the relationship between SST and hurricane numbers is *exponential* and poisson distributed.
In section 8 we derive expressions for aspects of the predicted distribution of landfalling hurricane numbers when predicted as a function of the number of basin hurricane numbers, in the case where the relationship between the two is linear and poisson distributed.
Finally in section 9 we discuss how to predict landfalling hurricane numbers from basin numbers which are themselves predicted from SST by combining the relationships derived in sections 5, 6 and 7 with the relationships derived in section 8.

2 Comments on our choice of models

This paper derives various mathematical expressions related to the prediction of hurricane numbers. It is not, however, a discussion of which models are actually appropriate to use to predict hurricane numbers. This latter question is a question we discuss at great length elsewhere: for a discussion of what models can be used to predict SST see Meagher and Jewson (2006) and Laepple et al. (2006); for a discussion of what models can be used to relate SST to landfalling hurricane numbers see Binter et al. (2006b); for a discussion of what models can be used to relate SST to basin hurricane numbers see Binter et al. (2006c); and for a discussion of what models can be used to relate basin hurricane numbers to landfalling hurricane numbers see Binter et al. (2006a).
However, we now give a brief summary of the rationale behind our model choices, based on the results from these studies:

- we consider a normal distribution for SST because that seems to be the simplest reasonable model
- we consider a linear-normal relationship between SST and hurricane numbers because this is the simplest case analytically, even though the use of the normal distribution for hurricane numbers may not be reasonable if the number of hurricanes is small
- we consider a linear-poisson relationship between SST and hurricane numbers because this is the simplest case that has a reasonable distribution for hurricane numbers even in the situation in which the number of hurricanes is small. We are not concerned that use of a linear relationship could in principle lead to a negative value for the poisson parameter since this does not happen in practice with the data we are using.
- we consider an exponential-poisson relationship between SST and hurricane numbers because this has been used previously (e.g. by Elsner and Schmertmann (1993)), and is the standard way that statisticians tend to use poisson regression. However, from the analysis we describe in Binter et al. (2006b) we conclude that it is not possible to tell from the data whether this model is better or worse than the linear-poisson model.
- we consider a linear-poisson relationship between basin hurricane numbers and landfalling hurricane numbers because it is simple, it does well in our own tests with real data (see Binter et al. (2006a)) and it includes the simpler model that consists of just a constant proportion.

3 Comments on notation and statistical models

Consider trying to build a statistical model that models some variables y_i as a function of some other variables x_i. One obvious place to start is standard linear regression, which can be written as:

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{1}$$

where the ϵ_i are taken to be IID and normally distributed with mean zero. From now on we drop the subscripts i for simplicity.
We can also write this model as:

$$y \sim N(\alpha + \beta x, \sigma^2) \tag{2}$$

In this paper we will call this the linear-normal model.
This model can be generalised in a number of ways. One way would be to change the distribution from normal to something else. We could do this either by specifying a different distribution for the noise

forcing ϵ, or by specifying a different distribution for the response y. In general these are not equivalent: for instance if we were to specify that the noise should be poisson distributed then y would have a poisson distribution shifted by $\alpha+\beta x$, and, conversely, if we were to specify that y should be poisson distributed then the noise would have a poisson distribution shifted by $-\alpha-\beta x$. In our case we think we have more idea about the distribution of y than we do about the distribution of the noise: since y is modelling hurricane numbers we think that it will be close to poisson. So the first generalisation we consider will be to replace equation 2 with:

$$y \sim \text{Po}(\text{rate} = \alpha + \beta x) \tag{3}$$

In other words: we model y as poisson distributed with mean $\alpha+\beta x$. We will call this the *linear-poisson* model.

We can also write this model as

$$y = \alpha + \beta x + \epsilon \tag{4}$$

just as before, but we note that although it is still the case that $E(\epsilon) = 0$, the distribution of ϵ is now somewhat odd, since it is a poisson distribution but shifted to have zero mean. As a result it is not common to write this model in this way (although there is nothing wrong with it).

We could also write this model using conditional expectations:

$$E(y|x) = \alpha + \beta x \tag{5}$$

with the additional information that y is poisson distributed.

We note that the linear poisson model does not often appear in statistics textbooks because for certain values of α, β and x the rate of the poisson could be negative, which is impossible. We, however, take a practical approach: for our data this is not a problem, and what happens far outside the domain covered by our data is not relevant to us and can be ignored.

Having discussed how to change the distribution used in standard linear regression, the second obvious generalisation would be to replace the linear function with something non-linear, and a common way to do this is to replace $\alpha+\beta x$ with $\exp(\alpha+\beta x)$. For those concerned about the problem with negative poisson rates described above, this solves that problem. We can write this new model as:

$$y \sim \text{Po}(\text{rate} = \exp(\alpha + \beta x)) \tag{6}$$

or

$$y = \exp(\alpha + \beta x) + \epsilon \tag{7}$$

or

$$E(y|x) = \exp(\alpha + \beta x) \tag{8}$$

where y is poisson distributed.

In fact, statisticians often write this model in yet another way, as:

$$\log E(y|x) = \alpha + \beta x \tag{9}$$

and call it the log-linear poisson regression model, although we find this nomenclature rather unhelpful, and we prefer to call this model an exponential-poisson regression model.

One final comment is that one can discuss all the models given above in the context of a general class of models known as generalised linear models (GLMs), which cover almost any possible distribution for y and any possible non-linear monotonic function of $\alpha+\beta x$.

4 Setup and basic relations

We will use the following notation:

- s is the SST
- s is taken as normal, with mean μ_s and sd σ_s
- n is the number of hurricanes
- $p()$ is used for probability densities

- $E()$ is used for expectations
- when we need to distinguish between the number of basin hurricanes and the number of landfalling hurricanes we write these as n_b and n_l.
- the relationships we use to model hurricane numbers as a function of SST give us the parameters α, β, σ
- the relationships we use to model landfalling hurricane numbers as a function of basin hurricane numbers give us the parameters α', β', σ'
- the mean and variance of the number of hurricanes is written as μ_h and σ_h^2, with μ_b and σ_b^2 for basin hurricanes and μ_l and σ_l^2 for landfalling

4.1 Calculating Means

We will calculate the mean number of hurricanes using:

$$\begin{aligned}
\mu_h &= E(n) && (10)\\
&= \sum_{n=0}^{\infty} np(n) && (11)\\
&= \sum_{n=0}^{\infty} n \int_{-\infty}^{\infty} p(n|s)p(s)ds && (12)\\
&= \int_{-\infty}^{\infty} \left(\sum_{n=0}^{\infty} np(n|s) \right) p(s)ds && (13)\\
&= \int_{-\infty}^{\infty} E(n|s)p(s)ds && (14)
\end{aligned}$$

4.2 Variances

We will calculate the variance of the number of hurricanes using:

$$\begin{aligned}
\sigma_h^2 &= E[(n-\mu_h)^2] && (15)\\
&= \sum_{n=0}^{\infty} (n-\mu_h)^2 p(n) && (16)\\
&= \sum_{n=0}^{\infty} (n-\mu_h)^2 \int_{-\infty}^{\infty} p(n|s)p(s)ds && (17)\\
&= \int_{-\infty}^{\infty} \left(\sum_{n=0}^{\infty} (n-\mu_h)^2 p(n|s) \right) p(s)ds && (18)\\
&= \int_{-\infty}^{\infty} E((n-\mu_h)^2|s)p(s)ds && (19)
\end{aligned}$$

In the non-linear cases we will calculate the variances using:

$$\begin{aligned}
sigma_h^2 &= E[(n-\mu_h)^2] && (20)\\
&= E(n^2) - \mu_h^2 && (21)
\end{aligned}$$

where

$$\begin{aligned}
E(n^2) &= \sum_{n=0}^{\infty} n^2 p(n) && (22)\\
&= \sum_{n=0}^{\infty} n^2 \int_{n=0}^{\infty} p(n|s)p(s)ds && (23)
\end{aligned}$$

$$= \int_{n=0}^{\infty} \left(\sum_{n=0}^{\infty} n^2 p(n|s) \right) p(s) ds \quad (24)$$

$$= \int_{n=0}^{\infty} E(n^2|s) p(s) ds \quad (25)$$

4.3 SST forecasts

We assume that we have some method for the prediction of SST that gives us a normal distribution $N(\mu_s, \sigma_s^2)$. The parameters μ_s and σ_s will typically have some uncertainty associated with them, which will depend on the prediction method being used. Some of the expressions we derive below need estimates of this uncertainty in order to derive estimates of the standard errors on our hurricane number predictions.

5 The linear-normal model for the relationship between SST and hurricane numbers

We now describe the first of our models for the relationship between SST and hurricane numbers, which is the linear-normal model. In this model we postulate a linear relation between SST s and hurricane numbers n, where n could be either the number of hurricanes in the basin, or the number at landfall. We assume that the distribution of hurricane numbers is normal. As discussed above, we include this model because it is perhaps the simplest model one might consider. However, the assumption of normality probably doesn't hold very well when the number of hurricanes is small: this is resolved in section 6 by replacing the normal distribution with the poisson.
The linear-normal model can be written as:

$$n = \alpha + \beta s + \epsilon \quad (26)$$

$$\epsilon \sim N(0, \sigma^2) \quad (27)$$

5.1 The predicted mean

Taking expectations of equation 26 (over all realisations of ϵ and s) gives us a simple expression for the mean number of predicted hurricanes in this model:

$$\mu_h = E(n) = \alpha + \beta E(s) = \alpha + \beta \mu_s \quad (28)$$

We see that the mean number of hurricanes is a linear function of the mean SST, and doesn't depend on the variance of SST σ_s^2.

5.2 The predicted variance

We can also derive an expression for the variance of the number of hurricanes σ_h^2 fairly easily.
Combining equations 26 and 28 we see that:

$$n - \mu_h = \beta(s - \mu_s) + \epsilon \quad (29)$$

$$(n - \mu_h)^2 = \beta^2 (s - \mu_s)^2 + 2\beta(s - \mu_s)\epsilon + \epsilon^2 \quad (30)$$

$$E((n - \mu_h)^2|s) = \beta^2 (s - \mu_s)^2 + \sigma^2 \quad (31)$$

where we take expecatations over all realisations of ϵ, for fixed s.
This gives:

$$\sigma_h^2 = \int E((n - \mu_h)^2|s) p(s) ds \quad (32)$$

$$= \int (\beta^2 (s - \mu_s)^2 + \sigma^2) p(s) ds \quad (33)$$

$$= \beta^2 \int (s - \mu_s)^2 p(s) ds + \int \sigma^2 p(s) ds \quad (34)$$

$$= \beta^2 \sigma_s^2 + \sigma^2 \quad (35)$$

This expression is easy to understand: the variance in the number of predicted hurricanes comes both from the variance in the SST prediction (scaled by β) and the variance around the relationship between mean SST and the mean number of hurricanes. The variance of the predicted number (b) the standard error on the SST prediction (which should be given by the SST prediction routine).

5.3 Standard errors

We now derive an approximate expression for the standard error on the prediction of the expected number of hurricanes.
We already have:

$$\mu_h = \alpha + \beta\mu_s \tag{36}$$

Now consider small errors in the parameters α and β and a small error in the prediction of the mean SST. We can understand what errors this leads to in our prediction of expected hurricane numbers just by linearising:

$$dm = d\alpha + \mu d\beta + \beta d\mu \tag{37}$$

Squaring this and taking expectations gives:

$$\text{var}(\mu_h) = \text{var}(\alpha_s) + \mu^2\text{var}(\beta) + \beta^2\text{var}(\mu_s) + 2\mu_s\text{cov}(\alpha,\beta) \tag{38}$$

This gives us an approximation to the standard error on the expected number of hurricanes in terms of the standard errors on the parameters from the regression (which are given by most regression routines), and the standard error on the SST prediction (which should be given by the SST prediction routine).

5.3.1 Higher order terms

By using Taylor expansions, we can derive higher order terms in the expression for the standard error. Consider the function $\mu_h = f(\alpha,\beta,\mu_s)$. Expanding this function in a Taylor series gives:

$$\begin{aligned}
d\mu_h &= \frac{\partial\mu_h}{\partial\alpha}d\alpha + \frac{\partial\mu_h}{\partial\beta}d\beta + \frac{\partial\mu_h}{\partial\mu_s}d\mu_s && (39)\\
&\quad +\frac{1}{2}\frac{\partial^2\mu_h}{\partial\alpha^2}d\alpha^2 + \frac{1}{2}\frac{\partial^2\mu_h}{\partial\beta^2}d\beta^2 + \frac{1}{2}\frac{\partial^2\mu_h}{\partial\mu_s^2}d\mu_s^2 && (40)\\
&\quad +\frac{\partial^2\mu_h}{\partial\alpha\partial\beta}d\alpha d\beta + \frac{\partial^2\mu_h}{\partial\beta\partial\mu_s}d\beta d\mu_s + \frac{\partial^2\mu_h}{\partial\mu_s\partial\alpha}d\mu_s d\alpha + ... && (41)\\
&= d\alpha + \mu_s d\beta + \beta d\mu_s && (42)\\
&\quad +0d\alpha^2 + 0d\beta^2 + 0d\mu_s^2 && (43)\\
&\quad +0d\alpha d\beta + 1d\beta d\mu_s + 0d\mu_s d\alpha + 0 && (44)\\
&= d\alpha + \mu_s d\beta + \beta d\mu_s + d\beta d\mu_s && (45)
\end{aligned}$$

This is exact, since all subsequent terms are zero. Squaring and taking expectations now gives:

$$\text{var}(d\mu_h) = \text{var}(\alpha) + \mu_s^2\text{var}(\beta) + 2\mu_s\text{cov}(\alpha,\beta) + \beta^2\text{var}(\mu_s) \tag{46}$$

and we find that our original expression is actually exact.

5.4 Linear sensitivity

We can also consider the linear sensitivity of our forecast to changes in the mean and the standard deviation of our SST prediction. This can be useful to understand how a change in the forecast will create a change in the hurricane prediction.
Again starting with

$$\mu_h = \alpha + \beta\mu_s \tag{47}$$

if we differentiate wrt μ_s we get:

$$\frac{\partial\mu_h}{\partial\mu_s} = \beta \tag{48}$$

and

$$\frac{\partial \mu_h}{\partial \sigma_s} = 0 \tag{49}$$

and we again see that the predicted mean number of hurricanes is independent of the variance in the SST forecast.

5.5 Summary

We now summarise the relations we have derived for the linear-normal model:

$$\begin{aligned}
\mu_h &= \alpha + \beta \mu_s && (50)\\
\sigma_h^2 &= \beta^2 \sigma_s^2 + \sigma^2 && (51)\\
\mathrm{var}(\mu_h) &= \mathrm{var}(\alpha) + \mu_s^2 \mathrm{var}(\beta) + \beta^2 \mathrm{var}(\mu_s) + 2\mu_s \mathrm{cov}(\alpha, \beta) && (52)\\
\frac{\partial \mu_h}{\partial \mu_s} &= \beta && (53)\\
\frac{\partial \mu_h}{\partial \sigma_s} &= 0 && (54)
\end{aligned}$$

We didn't actually use the fact that the SST was normally distributed to derive these relations: all we used were the values for the first two moments of the SST distribution. So all the above results hold for any SST distribution, given the first two moments.

5.6 Alternative representation

In practice, equation 52 is difficult to evaluate because the right hand side contains two large positive terms ($\mathrm{var}(\alpha)$ and $\mu_s^2 \mathrm{var}(\beta)$) and one large negative term ($2\mu_s \mathrm{cov}(\alpha, \beta)$). Rounding error can easily cause the result to be negative (when it should be positive), or at least very inaccurate.
We can avoid this problem by rewriting the original regression equation as:

$$n = \alpha + \beta(s - \overline{s}) + \epsilon \tag{55}$$

where $\overline{s}$ is the mean of the observed values of historical SST. Compared with the original formulation the value of β is the same but the value of α is now different.
This then gives:

$$\mu_h = \alpha + \beta(\mu_s - \overline{s}) \tag{56}$$

The expression for the variance, which doesn't depend on α, doesn't change.
The derivation for the standard errors is based on:

$$d\mu_h = d\alpha + (\mu_s - \overline{s})d\beta + \beta d\mu_s \tag{57}$$

giving:

$$\mathrm{var}(\mu_h) = \mathrm{var}(\alpha) + (\mu_s - \overline{s})^2 \mathrm{var}(\beta) + \beta^2 \mathrm{var}(\mu_s) \tag{58}$$

The $\mathrm{cov}(\alpha, \beta)$ term drops out because it is zero in this alternative representation.

6 The linear-poisson model for the relationship between SST and hurricane numbers

We now derive relations for the second of our models for the relationship between SST and hurricane numbers. This model is the same as the previous model, except that we replace the normal distribution with a poisson distribution. Because the expression for the mean number of hurricanes given the SST is still linear the results are rather similar.
We write this model as:

$$\begin{aligned}
n &= \alpha + \beta s + \epsilon && (59)\\
n &\sim \mathrm{Po}(\mathrm{rate} = \alpha + \beta s) && (60)
\end{aligned}$$

Note that the variance of n given s is given by $v(s) = \alpha + \beta s$ whereas in the previous model the variance was constant.

6.1 The predicted mean

Just as before:

$$\mu_h = E(n) = \alpha + \beta\mu_s \tag{61}$$

6.2 The predicted variance

$$\begin{aligned}
n - \mu_h &= \beta(s - \mu_s) + \epsilon && (62)\\
(n - \mu_h)^2 &= \beta^2(s - \mu_s)^2 + 2\beta(s - \mu_s)\epsilon + \epsilon^2 && (63)\\
E((n - \mu_h)^2|s) &= \beta^2(s - \mu_s)^2 + v(s) && (64)
\end{aligned}$$

where $v(s) = \alpha + \beta s$ and the expectation in the final step is again over realisations of the noise but not over realisations of the SST. The final equation is now slightly different than before because of the dependence of the variance on s. This has implications for the next step:

$$\begin{aligned}
\sigma_h^2 &= \int E((n - \mu_h)^2|s)p(s)ds && (65)\\
&= \int (\beta^2(s - \mu_s)^2 + v(s))p(s)ds && (66)\\
&= \beta^2 \int (s - \mu_s)^2 ds + \int v(s)p(s)ds && (67)\\
&= \beta^2 \int (s - \mu_s)^2 ds + \int (\alpha + \beta s)p(s)ds && (68)\\
&= \beta^2\sigma_s^2 + (\alpha + \beta\mu_s) && (69)\\
&= \beta^2\sigma_s^2 + \mu_h && (70)
\end{aligned}$$

We see that the part of the uncertainty in the hurricane number prediction that depends on the uncertainty in the SST-hurricane relationship now becomes μ_h, the predicted mean number of hurricanes. In other words, the higher the mean number of hurricanes predicted, the greater the uncertainty (in absolute terms).
Comparing the expression for the mean (equation 61) and the expression for the variance (equation 70) we see that, in general, they are not the same. We conclude that the predicted hurricane distribution, although a mixture of poisson distributions, is not itself a poisson distribution.

6.3 Standard errors

The derivation for the standard errors is the same as for the linear-normal case:

$$\mu_h = \alpha + \beta\mu_s \tag{71}$$

and so

$$d\mu_h = d\alpha + \mu_s d\beta + \beta d\mu_s \tag{72}$$

and

$$\mathrm{var}(\mu_h) = \mathrm{var}(\alpha) + \mu_s^2\mathrm{var}(\beta) + \beta^2\mathrm{var}(\mu_s) + 2\mu_s\mathrm{cov}(\alpha, \beta) \tag{73}$$

As before, this is in fact exact.

6.4 Linear sensitivity

This is also the same as for the linear-normal case:

$$\mu_h = \alpha + \beta\mu_s \tag{74}$$

so

$$\frac{\partial\mu_h}{\partial\mu_s} = \beta \tag{75}$$

and

$$\frac{\partial\mu_h}{\partial\sigma_s} = 0 \tag{76}$$

6.5 Summary

$$\begin{aligned}
\mu_h &= \alpha + \beta\mu_s && (77)\\
\sigma_h^2 &= \beta^2\sigma_s^2 + \alpha + \beta\mu_s && (78)\\
\mathrm{var}(\mu_h) &= \mathrm{var}(\alpha) + \mu_s^2\mathrm{var}(\beta) + \beta^2\mathrm{var}(\mu_s) + 2\mu_s\mathrm{cov}(\alpha,\beta) && (79)\\
\frac{\partial\mu_h}{\partial\mu_s} &= \beta && (80)\\
\frac{\partial\mu_h}{\partial\sigma_s} &= 0 && (81)
\end{aligned}$$

Once again the fact that we assumed SST was normal was irrelevant...and again all we used about the SST was the information about the first two moments.

6.6 Alternative representation

The alternative representation is almost the same as for the linear-normal case:

$$n = \alpha + \beta(s - \overline{s}) + \epsilon \tag{82}$$

This then gives:

$$\mu_h = \alpha + \beta(\mu_s - \overline{s}) \tag{83}$$

The derivation for the standard errors is based on:

$$d\mu_h = d\alpha + (\mu_s - \overline{s})d\beta + \beta d\mu_s \tag{84}$$

giving:

$$\mathrm{var}(\mu_h) = \mathrm{var}(\alpha) + (\mu_s - \overline{s})^2\mathrm{var}(\beta) + \beta^2\mathrm{var}(\mu_s) + 2(\mu_s - \overline{s})\mathrm{cov}(\alpha,\beta) \tag{85}$$

In this case the covariance term is not necessarily zero.

7 The exponential-poisson model for the relationship between SST and hurricane numbers

We now consider the third of our models for the relationship between SST and hurricane numbers, in which the mean number of hurricanes (given the SST) is given by an exponential of a linear function of SST, and the distribution of the number of hurricanes is poisson.

$$\begin{aligned}
n &= \exp(\alpha + \beta s) + \epsilon && (86)\\
n &\sim \mathrm{Po}(\mathrm{rate} = \exp(\alpha + \beta s)) && (87)
\end{aligned}$$

The variance depends on the SST, as for the linear-poisson model.
There are now big differences in the following analysis because of the non-linearity.

7.1 The predicted mean

Because of the non-linearity, evaluating the mean number of hurricanes is now a bit harder. We actually have to do the integral, and so we actually do have to use the assumption that the SST distribution is normal, rather than just needing the first two moments as in the previous two models.

$$\begin{aligned}
\mu_h &= \int \exp(\alpha + \beta s)p(s)ds && (88)\\
&= \int \exp(\alpha + \beta s)\frac{1}{\sqrt{2\pi}\sigma_s}\exp\left(-\frac{(s-\mu_s)^2}{2\sigma^2}\right)ds && (89)\\
&= \exp(\alpha)\frac{1}{\sqrt{2\pi}\sigma_s}\int \exp\left(\beta s - \frac{(s-\mu_s)^2}{2\sigma^2}\right)ds && (90)
\end{aligned}$$

$$= \exp(\alpha)\frac{1}{\sqrt{2\pi}\sigma_s}\int \exp\left[-\frac{1}{2\sigma_s^2}(-2\sigma_s^2\beta s + (s-\mu_s)^2\right] ds \quad (91)$$

$$= \exp(\alpha)\frac{1}{\sqrt{2\pi}\sigma}\int \exp\left[-\frac{1}{2\sigma_s^2}(s^2 - 2(\mu+\beta\sigma_s^2)s + \mu_s^2)\right] ds \quad (92)$$

$$= \exp(\alpha)\frac{1}{\sqrt{2\pi}\sigma_s}\int \exp\left[-\frac{1}{2\sigma_s^2}(s-(\mu+\beta\sigma_s^2))^2 - (\mu+\beta\sigma_s^2)^2 + \mu_s^2)\right] ds \quad (93)$$

$$= \exp(\alpha)\frac{1}{\sqrt{2\pi}\sigma_s}\int \exp\left[-\frac{1}{2\sigma_s^2}((s-a)^2 + b)\right] ds \quad (94)$$

$$= \exp\left(\alpha - \frac{b}{2\sigma_s^2}\right)\frac{1}{\sqrt{2\pi}\sigma_s}\int \exp\left[-\frac{1}{2\sigma_s^2}((s-a)^2)\right] ds \quad (95)$$

$$= \exp\left(\alpha - \frac{b}{2\sigma_s^2}\right)\frac{1}{\sqrt{2\pi}\sigma_s}\sqrt{2\pi}\sigma_s \quad (96)$$

$$= \exp\left(\alpha - \frac{b}{2\sigma_s^2}\right) \quad (97)$$

$$= \exp\left(\alpha + \beta(\mu_s + \beta\sigma_s^2/2)\right) \quad (98)$$

where we used the temporary variables:

$$a = \mu_s + \beta\sigma_s^2 \quad (99)$$

$$b = \mu_s^2 - a^2 \quad (100)$$

$$= \beta\sigma_s^2(2\mu_s - \beta\sigma_s^2) \quad (101)$$

7.2 The predicted variance

We can derive an expression for the variance in this case as follows:

$$\sigma_h^2 = E[(n-\mu_h)^2] \quad (102)$$

$$= E(n^2) - \mu_h^2 \quad (103)$$

μ_h^2 we know already, so we just need to calculate $E(n^2)$.

$$En^2 = \int E(n^2|s)p(s)ds \quad (104)$$

$$= \int E[\exp(\alpha+\beta s) + \epsilon]^2 p(s)ds \quad (105)$$

$$= \int [\exp(\alpha+\beta s)^2 + v(s)]p(s)ds \quad (106)$$

$$= \int [\exp(2\alpha+2\beta s) + v(s)]p(s)ds \quad (107)$$

$$= I_1 + I_2 \quad (108)$$

$$I_1 = \int \exp(2\alpha+2\beta s)p(s)ds \quad (109)$$

$$= \exp(2\alpha + 2\beta(\mu + 2\beta\sigma_s^2/2)) \quad (110)$$

$$= [\exp(\alpha + \beta(\mu + \beta\sigma_s^2/2))]^2 \exp(\beta^2\sigma_s^2) \quad (111)$$

$$= \mu_h^2 \exp(\beta^2\sigma_s^2) \quad (112)$$

$$I_2 = \int v(s)p(s)ds \quad (113)$$

$$= \int \exp(\alpha+\beta s)p(s)ds \quad (114)$$

$$= \mu_h \quad (115)$$

so

$$
\begin{aligned}
\sigma_h^2 &= E(n^2) - \mu_h^2 && (116)\\
&= \mu_h^2 \exp(\beta^2\sigma_s^2) + \mu_h - \mu_h^2 && (117)\\
&= \mu_h^2(\exp(\beta^2\sigma_s^2) - 1) + \mu_h && (118)\\
&\approx \mu_h^2\beta^2\sigma_s^2 + \mu_h && (119)
\end{aligned}
$$

where the final approximation will be accurate if $\beta^2\sigma_s^2 << 1$.

7.3 Standard errors

The derivation of expressions for standard errors follows the same logic as before, but is slightly more complicated because of the non-linearity:
The mean is given by:

$$
\begin{aligned}
\mu_h &= e^{\alpha+\beta(\mu_s+\beta\sigma_s^2/2)} && (120)\\
&= e^x && (121)
\end{aligned}
$$

where $x = \alpha + \beta(\mu_s + \beta\sigma_s^2/2)$.

$$
\begin{aligned}
d\mu_h &= \frac{\partial\mu_h}{\partial\alpha}d\alpha + \frac{\partial\mu_h}{\partial\beta}d\beta + \frac{\partial\mu_h}{\partial\mu_s}d\mu_s + \frac{\partial\mu_h}{\partial\sigma_s}d\sigma_s && (122)\\
&= e^x(d\alpha + (\mu_s + \beta\sigma_s^2)d\beta + \beta d\mu_s + \beta^2\sigma_s d\sigma_s) && (123)
\end{aligned}
$$

which gives:

$$
\begin{aligned}
\text{var}(\mu_h) &= e^{2x}[\text{var}(\alpha) + (\mu_s + \beta\sigma_s^2)\text{var}(\beta) + 2(\mu_s + \beta\sigma_s^2)\text{cov}(\alpha,\beta) && (124)\\
&\quad +\beta\text{var}(\mu_s) + \beta^2\sigma_s\text{var}(\sigma_s) + 2\beta^3\sigma_s\text{cov}(\mu_s,\sigma_s)] && (125)
\end{aligned}
$$

Unlike the linear cases this is now not exact, and there is in fact an infinite series of higher order terms. Evaluating the second order terms would be a useful way to assess the accuracy of the linear approximation.

7.4 Linear sensitivity

We can also calculate the linear sensitivity as follows:

$$
\begin{aligned}
\frac{\partial\mu_h}{\partial\mu_s} &= \frac{\partial}{\partial\mu_s}e^x && (126)\\
&= \beta e^x && (127)
\end{aligned}
$$

This makes sense: the sensitivity to the mean SST is proportional to the mean, and is proportional to β.
Also:

$$
\begin{aligned}
\frac{\partial\mu_h}{\partial\sigma_s} &= \frac{\partial}{\partial\mu_s}e^x && (128)\\
&= \beta^2\sigma_s e^x && (129)
\end{aligned}
$$

or

$$
\frac{1}{\sigma_s}\frac{\partial\mu_h}{\partial\sigma_s} = \beta^2 e^x \qquad (130)
$$

and we see that it is fractional changes in σ_s that matter. This is the only one of our three models where the standard deviation of the SST prediction σ_s has an impact on the mean hurricane numbers predicted. This arises because of the non-linearity: more uncertainty on the SST prediction leads to higher expected numbers of hurricanes.

7.5 Summary

$$\begin{aligned}
\mu_h &= \exp\left(\alpha+\beta(\mu_s+\beta\sigma_s^2/2)\right) && (131)\\
\sigma_h^2 &= \mu_h^2(\exp(\beta^2\sigma_s^2)-1)+\mu_h && (132)\\
\mathrm{var}(\mu_h) &= e^{2x}[\mathrm{var}(\alpha)+(\mu_s+\beta\sigma_s^2)\mathrm{var}(\beta)+2(\mu_s+\beta\sigma_s^2)\mathrm{cov}(\alpha,\beta) && (133)\\
&\quad +\beta\mathrm{var}(\mu_s)+\beta^2\sigma_s\mathrm{var}(\sigma_s)+2\beta^3\sigma_s\mathrm{cov}(\mu_s,\sigma_s)] && (134)\\
\frac{\partial\mu_h}{\partial\mu_s} &= \beta e^x && (135)\\
\frac{\partial\mu_h}{\partial\sigma_s} &= \beta^2\sigma_s e^x && (136)
\end{aligned}$$

7.6 Alternative representation

The alternative representation is also slightly different than before:

$$n=\exp(\alpha+\beta(s-\overline{s}))+\epsilon \tag{137}$$

This then gives:

$$\mu_h=\exp(\alpha+\beta(\mu_s-\overline{s}+\beta\sigma_s^2/2)) \tag{138}$$

The derivation for the standard errors is based on:

$$d\mu_h = e^x(d\alpha+(\mu_s-\overline{s}+\beta\sigma^2)d\beta+\beta d\mu_s+\beta^2\sigma_s d\sigma_s) \tag{139}$$

giving:

$$\begin{aligned}
\mathrm{var}(\mu_h) &= e^{2x}[\mathrm{var}(\alpha)+(\mu_s-\overline{s}+\beta\sigma_s^2)\mathrm{var}(\beta)+2(\mu_s-\overline{s}+\beta\sigma_s^2)\mathrm{cov}(\alpha,\beta) && (140)\\
&\quad +\beta\mathrm{var}(\mu_s)+\beta^2\sigma_s\mathrm{var}(\sigma_s)+2\beta^3\sigma_s\mathrm{cov}(\mu_s,\sigma_s)] && (141)
\end{aligned}$$

In this case the $\mathrm{cov}(\alpha,\beta)$ term doesn't disappear, but will be much smaller than in the original formulation.

8 The linear-poisson model for the relationship between basin and landfalling hurricane numbers

We now consider the relationship between the number of hurricanes in the basin and the number at landfall. We will use our model for this relationship later when we consider the possibility of predicting landfalling hurricane numbers in a three step approach: by first predicting SST, then predicting basin hurricane numbers from SST, and finally predicting landfalling hurricane numbers from basin hurricane numbers.

The only model we consider in this case is a linear-poisson model for the number of landfalling hurricanes as a function of the number of hurricanes in the basin:

$$\begin{aligned}
n_l &= \alpha'+\beta' n_b+\epsilon' && (142)\\
n_l &\sim \mathrm{Po}(\mathrm{rate}=\alpha'+\beta' n_b) && (143)
\end{aligned}$$

Comparing with the previous linear-poisson model described in section 6, the SST s has now become n_b. There are only slight differences in the analysis compared with that model.

8.1 The predicted mean

$$\mu_l=E(n)=\alpha'+\beta'\mu_b \tag{144}$$

8.2 The predicted variance

$$\begin{aligned} n - \mu_l &= \beta'(n_b - \mu_b) + \epsilon' & (145)\\ (n - \mu_l)^2 &= \beta'^2(n_b - \mu_b)^2 + 2\beta'(n_b - \mu_b)e + \epsilon'^2 & (146)\\ E((n - \mu_l)|n_b)^2 &= \beta'^2(n_b - \mu_b)^2 + v(n_b) & (147)\\ & & (148) \end{aligned}$$

where $v(n_b) = \alpha' + \beta' n_b$.

$$\begin{aligned} \sigma_l^2 &= \sum_{n=0}^{\infty} E((n - \mu_l)^2|n_b)p(n_b) & (149)\\ &= \sum_{n=0}^{\infty} (\beta'^2(n_b - \mu_b)^2 + v(n_b))p(n_b) & (150)\\ &= \beta'^2 \sum_{n=0}^{\infty} (n_b - \mu)^2 + \sum_{n=0}^{\infty} v(n_b)p(n_b) & (151)\\ &= \beta'^2\sigma_b^2 + \alpha' + \beta'\mu_b & (152)\\ &= \beta'^2\sigma_b^2 + \mu_l & (153) \end{aligned}$$

Once again, comparing the expression for the mean and the expression for the variance we see that, in general, they are not the same. The predicted hurricane distribution, being a mixture of poisson distributions, is not itself a poisson distribution.

8.3 Standard errors

$$\mu_l = \alpha' + \beta'\mu_b \tag{154}$$

so

$$d\mu_l = d\alpha' + \mu_b d\beta' + \beta' d\mu_b \tag{155}$$

and

$$\text{var}(\mu_l) = \text{var}(\alpha') + \mu_b^2\text{var}(\beta') + \beta'^2\text{var}(\mu_b) + 2\mu_b\text{cov}(\alpha', \beta') \tag{156}$$

As in the previous linear cases, this is exact.

8.4 Linear sensitivity

$$\mu_l = \alpha' + \beta'\mu_b \tag{157}$$

so

$$\frac{\partial\mu_l}{\partial\mu_b} = \beta' \tag{158}$$

It doesn't make sense to consider $\frac{\partial\mu_l}{\partial\sigma_b}$ since $\sigma_b^2 = \mu_b$.

8.5 Summary

$$\begin{aligned} \mu_l &= \alpha' + \beta'\mu_b & (159)\\ \sigma_l^2 &= \beta'^2\sigma_b^2 + \mu_l & (160)\\ \text{var}(\mu_l) &= \text{var}(\alpha') + \mu_b^2\text{var}(\beta') + \beta'^2\text{var}(\mu_b) + 2\mu_b\text{cov}(\alpha', \beta') & (161)\\ \frac{\partial\mu_l}{\partial\mu_b} &= \beta' & (162) \end{aligned}$$

8.6 Alternative representation

$$n_l = \alpha' + \beta'(n_b - \overline{n_b}) + \epsilon' \tag{163}$$
$$n_l \sim \text{Po}(\text{rate} = \alpha' + \beta'(n_b - \overline{n_b})) \tag{164}$$

$$\mu_l = \alpha' + \beta'(\mu_b - \overline{n_b}) \tag{165}$$

$$d\mu_l = d\alpha' + (\mu_b - \overline{n_b})d\beta' + \beta' d\mu_b \tag{166}$$

$$\text{var}(\mu_l) = \text{var}(\alpha') + (\mu_b - \overline{n_b})^2\text{var}(\beta') + \beta'^2\text{var}(\mu_b) + 2(\mu_b - \overline{n_b})\text{cov}(\alpha', \beta') \tag{167}$$

9 Predicting landfalls from basin numbers, and basin numbers from SST

The models given in sections 5, 6 and 7 for relating SST to hurricane numbers (which we now take as the number of hurricanes in the basin) can now be combined with the model given in section 8 that relates the numbers of hurricanes in the basin to the number at landfall.
We consider two cases below, based on the linear-poisson and exponential-poisson models for the number of hurricanes in the basin, and combined with the linear-poisson model for the number of hurricanes at landfall.

9.1 Linear poisson model from SST to basin, linear-poisson model from basin to landfall

9.1.1 Mean and variance

From section 6.5 we see that given an SST distribution $N(\mu_s, \sigma_s^2)$ the mean and variance of the distribution of the number of basin hurricanes is:

$$\mu_b = \alpha + \beta\mu_s \tag{168}$$
$$\sigma_b^2 = \beta^2\sigma_s^2 + \mu_b \tag{169}$$

and from section 8.5 we see that given the mean and variance of the number of basin hurricanes the mean and variance of the number of landfalling hurricanes is:

$$\mu_l = \alpha' + \beta'\mu_b \tag{170}$$
$$\sigma_l^2 = \beta'^2\sigma_b^2 + \mu_l \tag{171}$$

Putting these together in order to get from SST to landfalls in one step, we get:

$$\mu_l = \alpha' + \beta'\mu_b \tag{172}$$
$$= \alpha' + \beta'(\alpha + \beta\mu_s) \tag{173}$$
$$= \alpha' + \beta'\alpha + \beta'\beta\mu_s \tag{174}$$
$$\sigma_l^2 = \beta'^2\sigma_b^2 + \mu_l \tag{175}$$
$$= \beta'^2(\beta^2\sigma_s^2 + \mu_b) + \mu_l \tag{176}$$
$$= \beta'^2\beta^2\sigma_s^2 + \beta'^2\mu_b + \mu_l \tag{177}$$

9.1.2 Standard errors

For the SST to basin hurricanes part we have:

$$\text{var}(\mu_b) = \text{var}(\alpha) + \mu_s^2\text{var}(\beta) + \beta^2\text{var}(\mu_s) + 2\mu_s\text{cov}(\alpha, \beta) \tag{178}$$

and for the basin to landfall part we have:

$$\text{var}(\mu_l) = \text{var}(\alpha') + \mu_b^2\text{var}(\beta') + \beta'^2\text{var}(\mu_b) + 2\mu_b\text{cov}(\alpha', \beta') \tag{179}$$

and these two expressions can easily be combined to give a one-step expression for $\text{var}(\mu_l)$.

9.2 Exponential poisson model from SST to basin, linear-poisson model from basin to landfall

9.2.1 Mean and variance

From section 7.5 we see that given SST distribution $N(\mu_s, \sigma_s^2)$ the mean and variance of the distribution of the number of basin hurricanes is:

$$\mu_b = \exp(\alpha + \beta(\mu_s + \beta\sigma_s^2/2)) \quad (180)$$
$$\sigma_b^2 = \mu_b^2(\exp(\beta^2\sigma_s^2) - 1) + \mu_b \quad (181)$$

and from section 8.5 we again see that given the mean and variance of the number of basin hurricanes the mean and variance of the number of landfalling hurricanes is:

$$\mu_l = \alpha' + \beta'\mu_b \quad (182)$$
$$\sigma_l^2 = \beta'^2\sigma_b^2 + \mu_l \quad (183)$$

Putting these together:

$$\mu_l = \alpha' + \beta'\exp(\alpha + \beta(\mu_s + \beta\sigma_s^2/2)) \quad (184)$$
$$\sigma_l^2 = \beta'^2\mu_b^2(\exp(\beta^2\sigma_s^2) - 1) + \beta'^2\mu_h + \mu_l \quad (185)$$

9.2.2 Standard errors

For the SST to basin hurricanes part we have:

$$\mathrm{var}(\mu_b) = e^{2x}[\mathrm{var}(\alpha) + (\mu + \beta\sigma^2)\mathrm{var}(\beta) + \beta\mathrm{var}(\mu) + \beta^2\sigma\mathrm{var}(\sigma) + 2(\mu + \beta\sigma^2)\mathrm{cov}(\alpha, \beta)] \quad (186)$$

and for the basin to landfall part we have:

$$\mathrm{var}(\mu_l) = \mathrm{var}(\alpha') + \mu_b^2\mathrm{var}(\beta') + \beta'^2\mathrm{var}(\mu_b) + 2\mu_b\mathrm{cov}(\alpha', \beta') \quad (187)$$

and again these two expressions can easily be combined to give an expression for $\mathrm{var}(\mu_l)$.

References

R Binter, S Jewson, and S Khare. Statistical modelling of the relationship between hurricane numbers in the Atlantic Basin and at US landfall. *RMS Internal Report E05*, 2006a.

R Binter, S Jewson, and S Khare. Statistical modelling of the relationship between Main Development Region Sea Surface Temperature and Atlantic Basin hurricane numbers. *RMS Internal Report E04a*, 2006b.

R Binter, S Jewson, and S Khare. Statistical modelling of the relationship between Main Development Region Sea Surface Temperature and landfalling hurricane numbers. *RMS Internal Report E04b*, 2006c.

J Elsner and C Schmertmann. Improving extended-range seasonal predictions of intense Atlantic hurricane activity. *Weather and Forecasting*, 3:345–351, 1993.

T Laepple, S Jewson, J Meagher, A O'Shay, and J Penzer. Five-year ahead prediction of Sea Surface Temperature in the Tropical Atlantic: a comparison of simple statistical methods. *arXiv:physics/0701162*, 2006.

J Meagher and S Jewson. Year ahead prediction of hurricane season SST in the tropical Atlantic. *arxiv:physics/0606185*, 2006.